Structure of the Brd4 ET domain bound to a C-terminal motif from γ-retroviral integrases reveals a conserved mechanism of interaction

Brandon L. Crowe[a], Ross C. Larue[b], Chunhua Yuan[c], Sonja Hess[d], Mamuka Kvaratskhelia[b], and Mark P. Foster[a,1]

[a]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210; [b]Center for Retrovirus Research and College of Pharmacy, The Ohio State University, Columbus, OH 43210; [c]Campus Chemical Instrument Center, The Ohio State University, Columbus, OH 43210; and [d]Proteome Exploration Laboratory, California Institute of Technology, Pasadena, CA 91125

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved January 14, 2016 (received for review August 24, 2015)

The bromodomain and extraterminal domain (BET) protein family are promising therapeutic targets for a range of diseases linked to transcriptional activation, cancer, viral latency, and viral integration. Tandem bromodomains selectively tether BET proteins to chromatin by engaging cognate acetylated histone marks, and the extraterminal (ET) domain is the focal point for recruiting a range of cellular and viral proteins. BET proteins guide γ-retroviral integration to transcription start sites and enhancers through bimodal interaction with chromatin and the γ-retroviral integrase (IN). We report the NMR-derived solution structure of the Brd4 ET domain bound to a conserved peptide sequence from the C terminus of murine leukemia virus (MLV) IN. The complex reveals a protein–protein interaction governed by the binding-coupled folding of disordered regions in both interacting partners to form a well-structured intermolecular three-stranded β sheet. In addition, we show that a peptide comprising the ET binding motif (EBM) of MLV IN can disrupt the cognate interaction of Brd4 with NSD3, and that substitutions of Brd4 ET residues essential for binding MLV IN also impair interaction of Brd4 with a number of cellular partners involved in transcriptional regulation and chromatin remodeling. This suggests that γ-retroviruses have evolved the EBM to mimic a cognate interaction motif to achieve effective integration in host chromatin. Collectively, our findings identify key structural features of the ET domain of Brd4 that allow for interactions with both cellular and viral proteins.

NMR | protein–protein interaction | BET family transcription factor | retroviral integration | ET binding motif (EBM)

The bromodomain and extraterminal domain (BET) family of proteins (Brd2, 3, 4, and T) play central roles in regulating cell fate (1). They have been implicated in diverse cellular phenomena, including inflammation, obesity, spermatogenesis, cancer, DNA damage repair, viral latency, and, more recently, γ-retroviral integration site selection (2–8). Most of these functions have been associated with its dual role as epigenetic reader and transcriptional activator (1–4).

The BET family of proteins, as well as the extended BET family (Brd1, 7, 8, and 9), is characterized by two N-terminal bromodomains and the extraterminal (ET) domain, for which the proteins are named; Brd4 is known to occur in two different isoforms, with the longer variant containing a C-terminal motif (9) (Fig. 1*A*). The dual bromodomains recognize and bind acetylated lysines on histone H3 and H4 tails on chromatin (9–11), thereby tethering the protein and any associated factors to those histone posttranslational modifications (PTMs). Whereas recognition of select histone marks is achieved through the bromodomains, the overall high-affinity binding to chromatin is the result of cooperative binding to both its cognate PTMs and nonspecific binding to DNA wrapped around mononucleosomes through interaction with the conserved motifs A and B (12, 13). The C-terminal domain of the long Brd4 isoform functions in activation of RNA polymerase II via interactions with pTEFb (positive transcription elongation factor) and cooperates with the jumonji C-domain-containing protein 6 (JMJD6) in regulating release of promoter-proximal pausing (14).

Brd4 also exhibits a range of pTEFb-independent functions. Isoforms lacking the C-terminal motif have been shown to function in DNA damage repair via interactions with components of the condensin II chromatin remodeling complex (8). Proteomics and knockdown experiments demonstrated that the ET domain mediates transcriptional activation through recruitment of several chromatin-modifying enzymes, including JMJD6, histone-lysine N-methyltransferase NSD3 (NSD3), glioma tumor suppressor candidate region gene 1 protein (GLTSCR1), ATPase family AAA domain-containing protein 5 (ATAD5), and chromodomain helicase DNA-binding protein 4 (CHD4) (15). The interaction with NSD3 underlies an aggressive midline carcinoma resulting from a chromosomal translocation that fuses Brd4 with nuclear protein in testes (16). Owing to their role in cell cycle regulation and epigenetic sensing, as well as associations with a range of disease states, the therapeutic potential of targeting BET proteins with small molecule inhibitors is an area of intense interest, most of which is centered on preventing chromatin binding by the bromodomains (10, 11).

Significance

Bromodomain and extraterminal domain (BET) proteins recruit a variety of cellular factors to epigenetic marks on chromatin to modulate essential cellular processes, including DNA repair and transcription. To facilitate their recruitment to chromosomal DNA made accessible by epigenetically modified chromatin dynamics, conserved sequences in γ-retroviral integrases bind to the highly conserved extraterminal (ET) domains of BET proteins. The solution structure of the complex between the Brd4 ET domain and a conserved sequence from murine leukemia virus integrase reveals the basis for γ-retroviral integration pattern, with implications for retrovirus-based gene therapy. Moreover, the studies suggest how BET proteins might interact with cognate binding partners involved in a diverse array of cellular phenomena including transcriptional activation, cancer, viral latency, and viral integration.

Author contributions: B.L.C., R.C.L., C.Y., M.K., and M.P.F. designed research; B.L.C., R.C.L., C.Y., S.H., and M.P.F. performed research; B.L.C., R.C.L., C.Y., S.H., M.K., and M.P.F. analyzed data; and B.L.C., R.C.L., C.Y., S.H., M.K., and M.P.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: NMR, atomic coordinates, chemical shifts, and restraints have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2N3K), and BioMagResBank, www.bmrb.wisc.edu (BMRB ID code 25649).

[1]To whom correspondence should be addressed. Email: foster.281@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516813113/-/DCSupplemental.

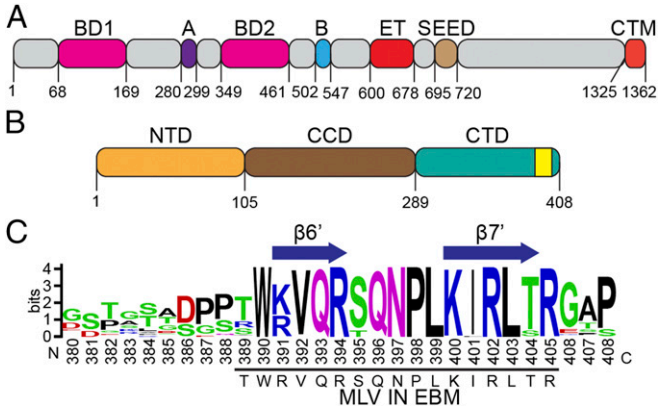

Fig. 1. Domain maps of Brd4 (*A*) and MLV integrase (*B*); the ET-binding motif (EBM) is at the extreme C terminus (yellow). (*C*) Weblogo (weblogo.berkeley.edu/) of sequence conservation at the C terminus of γ-retroviral integrases, with the EBM peptide used in these studies indicated below; residues 390–405 are highly conserved. β strands formed upon binding the ET domain are indicated by block arrows (see text).

In addition to recruiting chromatin host factors involved in transcriptional regulation, the ET domain has been implicated as the binding site for viral factors, including γ-2 herpesvirus latency-associated nuclear antigen and murine leukemia virus (MLV) integrase (IN) (12, 17–19). Analyses of MLV integration site selection revealed preferential integration at regions enriched in acetylated histone PTMs, such as transcription start sites (TSS) and enhancers, implicating guiding by BET proteins through bimodal interactions with nucleosomes and MLV IN (6, 12, 20–22). Although solution structures of the isolated MLV IN C-terminal domain and of the ET domain have been reported, the structural determinations of their interaction have not. Mutagenesis experiments have shown that a conserved sequence motif from the extreme C terminus of MLV IN is responsible for specific tethering to the ET domain (Fig. 1*B*, yellow) (12, 23), and NMR studies of the MLV IN C-terminal domain revealed that this region is unstructured, suggesting a "fly-casting" mechanism for BET protein binding (23). Modifications to this region compromise MLV IN binding to BET proteins and alter MLV integration patterns (12, 23–25). In addition, the sequence of the ET domain is highly conserved (Fig. S1), adopting a structure that features three α helices connected by two loops (26), whereas its mode of interaction with other protein factors has yet to be established.

We determined the solution structure of a complex between the Brd4 ET domain and a peptide comprising the conserved C-terminal sequence from MLV IN. Upon binding, both protein and peptide undergo binding-coupled folding to form an intermolecular β sheet and make an array of specific side-chain interactions. We show that these features are important for binding of host transcription factors to full-length Brd4, and that the peptide serves as an effective inhibitor of at least some of these interactions.

Results

Binding of ET Binding Motif to Brd4 ET. A peptide comprising the 17 conserved residues from MLV IN (Fig. 1*C*) is necessary and sufficient for ET domain binding, with an apparent affinity of ∼160 nM as measured by isothermal titration calorimetry (ITC) (Fig. 2*A*). NMR titrations show that the ET binding motif (EBM) binds in the slow-exchange regime, and that the interaction is coupled to folding of both partners, as evident from large chemical shift perturbations and increase in resonance dispersion (Figs. 2 *B* and *C* and 3*B* and Fig. S2). Furthermore, significant downfield chemical

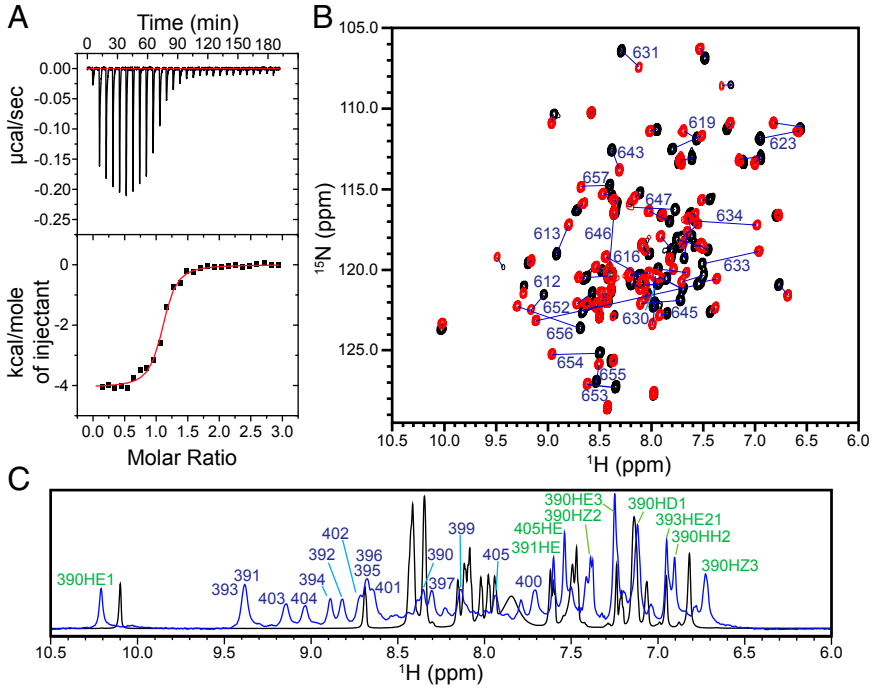

Fig. 2. Induced-fit binding between Brd4 ET domain and the MLV IN EBM. (*A*) Isothermal titration calorimetry of the ET–EBM interaction reveals high affinity, with K_d 159 ± 12 nM. (*B*) ^{1}H-^{15}N–correlation spectra of [U-^{15}N]-Brd4 ET domain in the absence (black) and presence of unlabeled MLV IN EBM (red). Resonances with large chemical shift perturbations (CSPs) are labeled with their amino acid assignment. (All backbone amide assignments and plot of CSPs are shown in Fig. S2.) (*C*) One-dimensional ^{1}H spectrum of the MLV EBM free (black) and bound to the ET domain, with protein signals suppressed by a [^{13}C, ^{15}N] filter (blue). Peaks labels indicate well-dispersed signals in the bound peptide, blue labels for backbone signals and green for side-chain signals.

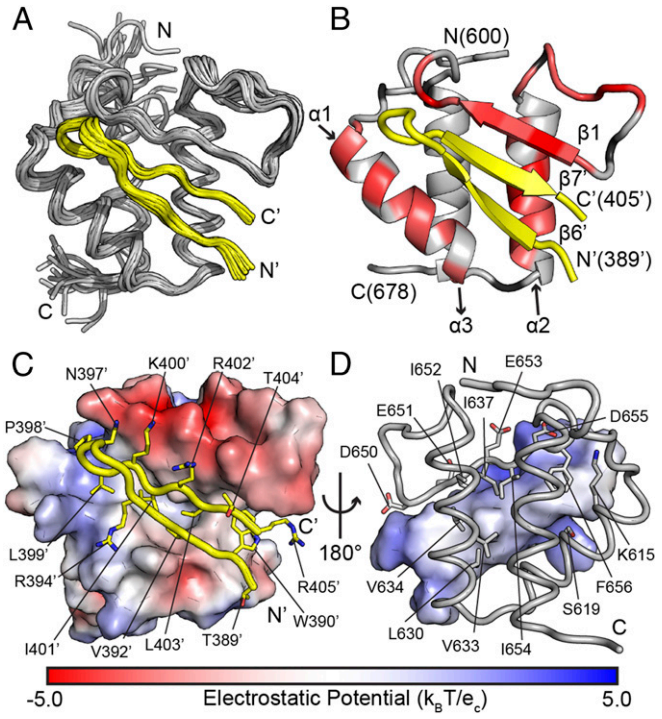

Fig. 3. Solution structure of the complex between Brd4 ET and MLV IN EBM. (*A*) Ensemble of 20 lowest-energy structures of the complex between the Brd4 ET domain (gray) and EBM (yellow). (*B*) Ribbon diagram of the lowest-energy conformer; the new β sheet comprises antiparallel strands β1 from the ET domain (residues 650–654) and β6′ (391′–395′) and β7′ (400′–404′) from the EBM (Fig. S3). The protein ribbon is colored according to backbone amide shift perturbations (as quantified in Fig. S2). (*C* and *D*) Electrostatic potential map of the binding interface between Brd4 ET and the EBM, illustrating charged and hydrophobic residues in the intermolecular interface (Fig. 4).

shifts for several H^{α} spins in both the ET and EBM indicate formation of a β sheet.

Solution Structure of the ET Domain–EBM Complex. We determined the solution structure of [U-^{13}C, ^{15}N]-labeled Brd4 ET (600–678) bound to unlabeled MLV integrase EBM (389–405) from distance and torsion angle restraints. Complete Brd4 ET backbone assignments and 86% side-chain assignments were obtained using standard triple-resonance NMR spectra. Most of the missing side-chain assignments are due to spectral overlap of lysine resonances. The unlabeled peptide ^{1}H assignments were achieved to 88% using 2D ^{13}C/^{15}N-filtered COSY, total correlation spectroscopy (TOCSY), and NOESY (Fig. S3*A*); 817 distance restraints for Brd4 ET were obtained from 3D ^{13}C-edited or ^{15}N-edited NOESY spectra, 174 distance restraints for the MLV IN EBM were obtained from a 2D ^{13}C/^{15}N-filtered NOESY spectrum, and 277 intermolecular distance restraints were obtained from 3D ^{13}C/^{15}N-filtered (*f1*), ^{13}C-edited, or ^{15}N-edited (*f3*) NOESY spectra. Torsion angle restraints were obtained from chemical shift data using TALOS-N (Table S1).

The solution structure was well-defined with an all-heavy-atom rmsd of 1.86 ± 0.20 Å for the complex (Fig. 3*A*). The Brd4 ET domain retained three α helices observed in the spectra of the free Brd4 ET domain. All helices displayed NOE patterns consistent with helices including midrange αN(i,i+3), NN(i,i+2), and αβ(i,i+3) NOEs. Helix α1 spans residues Y612 to K624, helix α2 spans residues G627 to R640, and helix α3 spans P661 to L675 (Fig. 3*B*).

The interface between the Brd4 ET domain and the MLV IN peptide features a three-stranded antiparallel β sheet and contacts involving helices α1 and α2, burying ~780 Å^2 of solvent-accessible surface area on the ET domain. The newly formed β sheet consists of one strand (β1) from the loop between helices α2 and α3 of the ET domains and two strands from the MLV IN peptide folding to form a β-hairpin (β6′, β7′); the sheet is well-defined by NOEs characteristic of β–strand formation (Fig. S3*B*). Interestingly, both the Brd4 ET domain loop region and the C-terminal tail of MLV integrase are poorly structured on their own (23, 26). Thus, in the context of the isolated domains, induced fit folding of both components is required to form the interface, which partly explains the large chemical shift perturbations observed within this region of the Brd4 ET domain (Fig. S2). The interface features both hydrophobic and electrostatic interactions on opposite faces of the β sheet (Fig. 3 *C* and *D*).

The structural features in the ET–EBM interface provide a strong rationale for the observed sequence conservation of the extreme C terminus of the γ-retroviral integrases, mirroring the high degree of sequence conservation of the ET domains. Paired strands β1 and β7′ are characterized by a pattern of alternating hydrophobic and charged amino acids. The exterior surface of the sheet features complementary electrostatic interactions between the negatively charged side chains on the conserved DEIDIDF(650–656) of the ET domain β1 and the positively

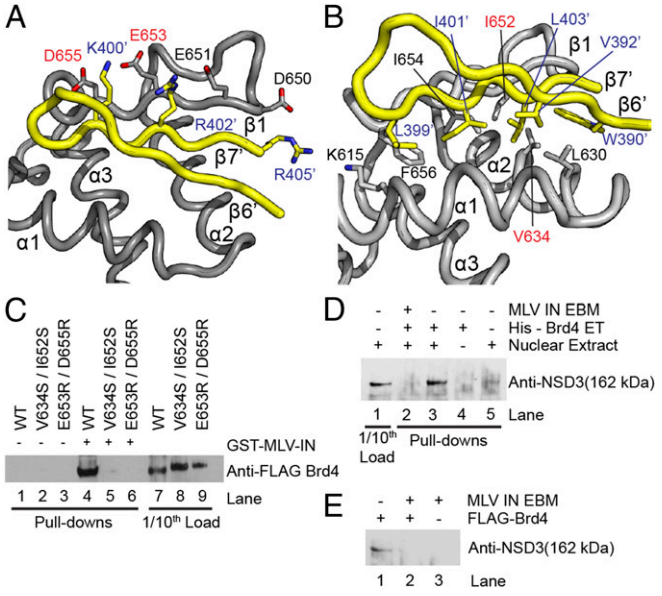

Fig. 4. Specific side-chain interactions between Brd4 ET and MLV IN EBM. (*A*) Complementary arrangement of positively charged side chains on β7′ of IN (blue) and negatively charged side chains of the ET domain (black); ET domain residues mutated for binding studies are labeled in red. (*B*) Hydrophobic interface defined by conserved nonpolar side chains on Brd4 and MLV IN. Label colors as in *A*. (*C*) Affinity capture assay shows that MLV IN binds to WT FLAG-tagged Brd4 (1–720) (lane 4), but not mutants designed to disrupt the ET–EBM interface, V634S/I652S and E653R/D655R (lanes 5 and 6). No Brd4 is detected bound to the resin in the absence of GST-tagged MLV-IN (lanes 1–3); lanes 7–9 are immunoblots of 1/10th of the protein loaded in lanes 1–6. (*D*) The His-tagged Brd4 ET domain can bind NSD3 in the absence (lane 3), but not in the presence, of the EBM peptide (lane 2), demonstrating competition between the viral and host factors for ET binding. Lane 1 shows the presence of NSD3 in the extract, lane 4 shows no cross-reactivity, and lane 5 shows that NSD3 is not retained on the Ni resin in the absence of Brd4. (*E*) The FLAG-tagged Brd4 can bind NSD3 in the absence (lane 1), but not in the presence, of the EBM peptide (lane 2), demonstrating competition between the viral and host factors for ET binding. Lane 3 shows that NSD3 is not retained on the Ni resin in the absence of Brd4.

charged side chains on the conserved LKIRLTR(399′–405′) on the EBM β7′ (Fig. 4*A*). Additionally, the buried side of the sheet features favorable hydrophobic interactions involving residues contributed by helices α1 and α2 and the strand β1 of the ET domain (L630, V634, I652, I654, and F656) and both strands of the EBM (W390′, V392′, L399′, I401′, and L403′) (Fig. 4*B*). These structural features are consistent with previous biochemical and mutagenesis studies. For example, mutation to Ala of MLV IN W390′, which is positioned directly below strand β6′ and makes hydrophobic contacts to L630 and V634, shifts the integration profile of MLV-based vectors away from TSS (24). Likewise, affinity capture experiments using MLV IN and Brd2 mutants showed that ET residues corresponding to Brd4 ET L630, D655, F656, and E666 (Brd2 residues L662, D687, F688, and E657) are critical for interaction with γ-retroviral integrases (18).

ET Domain Residues Recognized by the EBM Are Also Important for Brd4 Binding to Other Cellular Components. The importance of observed intermolecular contacts in molecular recognition by full-length Brd4 were validated by site-directed mutagenesis: In affinity capture experiments, Brd4 E653R/D655R and V634S/I652S mutants fail to bind MLV IN (Fig. 4*C*, lanes 4–6), consistent with published findings with Brd2 (18). We also tested the effect of these mutations on binding to host factors in 293T cells by affinity capture and proteomic analysis of ectopically expressed FLAG-tagged WT and mutant Brd4 (Fig. S4 and Dataset S1). The proteomic data do not distinguish between direct and indirect interactions, and overall peptide counts were observed to decrease between WT Brd4 and the mutants; however, proteins previously identified as ET domain-interacting proteins, including NSD3 (and related NSD2), ATAD5, GLTSCR1, and CHD4 (and related CHD7) (14, 15), could be identified as binding partners of WT Brd4 but neither mutant. In addition, similar profiles were observed for many proteins not already known to bind Brd4 (as itemized in thebiogrid.org/), consistent with the participation of Brd4 in interactions with a variety of higher-order protein complexes. Importantly, interactions with histone proteins were largely unaffected by the mutations, indicating that mutations to the ET domain do not prevent Brd4 binding to chromatin via its N-terminal bromodomains. We conclude that the structural features on the Brd4 ET domain that are recognized by MLV IN are also important for host-factor binding, which suggest a structural basis for the participation of BET proteins in a variety of chromatin-centered transactions.

The EBM Peptide Competes with NSD3 for the Brd4 ET Domain. A peptide comprising the MLV IN-derived EBM can displace interactions with host factors that target the Brd4 ET domain. Using affinity capture and antibody detection, we found that NSD3 binds either the full-length Brd4 or the isolated ET domain in the absence of the peptide, whereas 10 μM EBM abolishes the interaction (Fig. 4*D*, lanes 2 and 3, and Fig. 4*E*, lanes 1 and 2). This finding supports the hypothesis that the structural features conserved in the ET domain-binding region of the γ-retroviral integrases are representative of a more universal mechanism of host-factor recruitment to chromatin by BET proteins. Thus, the compact ET binding motif defined by the C terminus of the MLV IN may comprise the key features of ET domain recognition by both viral and host factors.

Discussion

We report the solution structure of a high-affinity complex formed between the ET domain of Brd4 and the EBM, a peptide comprising the interaction region from MLV IN. The NMR studies reveal that these components interact via binding-coupled folding, in which regions that were poorly structured in isolation (23, 26) fold to form a new three-stranded β sheet that generates complementary pairing of charged and hydrophobic residues on both molecules. In cells, interactions between Brd4 and MLV IN are critical for the recruitment of the MLV preintegration complex (PIC) to host chromatin during integration. The PIC is composed of a multimer of IN stably bound to viral DNA ends with additional viral and cellular proteins (27, 28), which then engages with Brd4 and its associated cognate chromatin binding partners (4). Although the NMR data do not directly inform whether the interacting regions of Brd4 and IN are unstructured in the context of the intact PIC, amino acid substitutions at the ET–EBM interface abolish the interaction between the full-length recombinant Brd4 and MLV IN proteins (Fig. 4*C*) and affect MLV integration (18, 24). This suggests that the intricacy and complementarity of the interface revealed by the NMR structures is responsible for their tight and specific binding.

The role of BET proteins in guiding MLV integration site selection to TSS and enhancers through the recognition of acetylated histone tails is comparable to the role of LEDGF/p75 guiding lentiviral integration toward active transcription units through the H3K36me3 epigenetic mark (29–35). Although both cellular proteins use a bimodal mechanism for interaction with chromatin and retroviral IN, the structural details of these interactions are very different. In the case of lentiviruses and LEDGF/p75, a short interhelical loop of the integrase binding domain (IBD) of LEDGF/p75 engages a cavity on the dimer interface of two HIV-1 catalytic core domains (35, 36). The essential interacting features of LEDGF/IBD and HIV-1 IN are not conserved in the interaction between BET proteins and MLV IN. Instead, we observe a tight interaction governed by the formation of an intermolecular three-stranded β sheet upon binding, with specificity governed by a pattern of alternating hydrophobic and charged amino acids on both interfaces.

These studies have implications for the use of MLV-based vectors in gene therapy. MLV-based vectors have been curatively used in human gene therapy for treatment of several genetic deficiencies, including X-linked severe combined immune deficiency, Wiscott–Aldrich syndrome, and X-linked chronic granulomatous disease (36–40). However, a common feature of all of these studies is that a significant number of patients subsequently developed leukemia due to transcriptional up-regulation of proto-oncogenes (41–43). Because BET proteins are the primary determinants of MLV integration-site selection (6), structural characterization of the ET–EBM complex allows us to better understand the mechanisms of γ-retroviral integration site selectivity.

The observation that EBM residues important for ET binding are highly conserved (Fig. 1) and interact with highly conserved sequences in the ET domains (Figs. 3 and 4 and Fig. S1) suggests that γ-retroviruses have evolved to mimic a conserved native protein–protein interaction through modification of its C terminus, which is not essential for catalytic activity in vitro or in cells (12, 37). Indeed, our site-directed mutagenesis and proteomic experiments indicate that the structural features of Brd4 important for binding by MLV IN are also important for binding to cellular proteins previously identified to interact with the ET domain, including NSD3, ATAD5, GLTSCR1, and CHD4 (Dataset S1), suggesting common interaction determinants. A similar theme of structural mimicry of chromatin-associated proteins by integrases is observed in yeast, wherein retrotransposon Ty5 IN guides integration to regions of silent chromatin via interactions between its C terminus with the heterochromatin protein Sir4, using the same site as that of the nuclear membrane associated protein Esc1, although the structural details differ between Ty5 and MLV IN (38, 39).

Our structural and binding studies suggest a means for interfering with specific aspects of BET protein function. BET proteins, and Brd4 in particular, have attracted intense interest because of their pleiotropic functions in a range of human diseases (40). Most of this effort has been focused on selectively targeting the bromodomains of BET proteins with mimics of

acetylated lysine to prevent their association with chromatin (41–43). We show that a peptide comprising the EBM can compete with the transcription factor NSD3 for binding to full-length Brd4, highlighting the importance of the ET domain in the interaction. Considering the compactness of the structural motif represented by the EBM, and its high-affinity binding to the ET domain, we imagine that small molecule ligands modeled after it might find use for disrupting interactions between Brd4 and a subset of its interacting partners implicated in cellular phenomena that include transcriptional activation, cancer, viral latency, and viral integration.

Experimental Procedures

Plasmids and Protein Purification. C-terminal FLAG-tagged human full-length Brd4(1–720) (amino acids 1–720; NM_014299) were purchased from Origene (6). Mutations in full-length FLAG-Brd4(1–720) were introduced using a site-directed mutagenesis kit (Stratagene). His-Brd4(1–720) was generated and purified as previously described (6). HEK293T cells were cultured in Dulbecco's modified eagle medium (Invitrogen), 10% (vol/vol) FBS (Invitrogen), and 1% (vol/vol) antibiotic (Gibco) at 37 °C and 5% (vol/vol) CO_2. All ectopic expression plasmids were transfected into HEK293T cells using TransIT-2020 (Mirus). GST-MLV IN (amino acids 1–408) was generated and purified as previously described (6).

Expression and Purification of Recombinant Isotopically Labeled Brd4 ET. The Brd4 ET domain was expressed in *Escherichia coli* BL21(DE3) cells (Agilent) grown in M9 minimal media supplemented with 1% (vol/vol) Eagle Basal Vitamin Mix (Life Technologies, Gaithersburg, MD), containing 1 g/L ^{15}N-ammonium chloride (Cambridge Isotope Laboratories, Inc.) and 2 g/L ^{13}C-glucose (Cambridge Isotope Laboratories, Inc.) as sole nitrogen and carbon sources, respectively. Cultures were grown at 37 °C to an optical density of 0.70 at 600 nm and induced for 16 h at 37 °C by addition of 0.5 mM isopropyl-thio-β-D-galactopyranoside. Bacterial cells from 1 L of culture were disrupted by sonication in buffer containing 20 mM sodium phosphate, pH 7.0, 500 mM NaCl, 15 mM imidazole, one-half tablet of the complete EDTA-free protease inhibitor (Roche), and 50 μM PMSF (Thermo Scientific). The protein was purified using a 5 mL HisTrap HP (GE Healthcare) using a gradient from 15 mM to 500 mM imidazole in 20 mM sodium phosphate, pH 7.0. The N-terminal hexahistidine tag was removed by tobacco etch virus (TEV) protease digestion at 25 °C for 16 h, leaving five nonnative amino acids (Gly-Ala-Ile-Ala-Met) on the N terminus and two nonnative amino acids (Thr-Arg) on the C terminus. The protein was the purified from the hexahistidine tag fragment and the TEV protease by a second pass through the 5-mL HisTrap HP column, this time collecting the flow-through. Finally, the protein was purified by size-exclusion chromatography using HiLoad 16/60 Superdex 75 (GE Healthcare) in 20 mM sodium phosphate, pH 7.0, 500 mM NaCl, and 5 mM DTT. Purified Brd4 ET was exchanged into NMR buffer (20 mM d_{11}-Tris, pH 7.0, 150 mM NaCl, and 5 mM d_6-DTT) using a PD10 column (GE Healthcare).

Isothermal Titration Calorimetry. ITC was performed by titrating MLV IN EBM (206 μM) into Brd4 ET (14.9 μM) in buffer containing 100 mM NaCl, 1 mM DTT, and 20 mM Tris at pH 7.0. The titration was performed using a MicroCal VP-ITC at 25 °C with 28 injections (5 μL for the first injection and 10 μL for all subsequent injections) with 400 s between injections. The first injection was discarded and the data were fit to a one-site model using the Origin version 7 software package (MicroCal). Best-fitted parameters were $n = 1.07 \pm 0.01$, $\Delta H = -4{,}040 \pm 43.6$ cal/mol, and $K_d = 159 \pm 12$ nM.

NMR Structure Determination. [U-^{15}N]- or [U-^{15}N,^{13}C]- Brd4 ET was concentrated to 0.4 mM in buffer containing 100 mM NaCl, 5 mM d_6-DTT, and 20 mM d_{11}-Tris, pH 7.0. D_2O was added to 10% (vol/vol) and DSS to 0.66 mM to the NMR sample. Unlabeled peptide containing residues 389–405 of MLV integrase (the EBM), obtained from Biomatrik, Inc., was suspended in water to make 2 mM stock solution and then titrated into NMR samples. The formation of the high-affinity protein–peptide complex, which is in slow exchange on the NMR timescale, was monitored by 2D ^{1}H-^{15}N–correlated spectra of the labeled Brd4 ET domain as well as by 1D ^{13}C/^{15}N-filtered ^{1}H spectra of the unlabeled peptide.

Structural determination followed a well-established protocol using multidimensional and triple-resonance experiments (44–46). All of the datasets were recorded at 25 °C on Bruker Avance DRX-600, DRX-700, and DRX-800 spectrometers equipped with a 5-mm triple-resonance cryoprobe and *z* axis gradient. Data were processed with NMRPipe (47) and visualized/analyzed with NMRViewJ (48).

For protein backbone assignments, 3D triple-resonance spectra HNCO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH (44) were recorded on purified [U-^{13}C, ^{15}N] Brd4 ET containing the uncleaved hexahistidine tag and TEV protease site in complex with the MLV IN EBM. In addition, HNCA and HNCO spectra were recorded on the complex of [U-^{13}C, ^{15}N] Brd4 ET after the removal of the hexahistidine tag, bound to unlabeled MLV IN EBM: Only plasmid-encoded residues adjacent to the linker exhibited differences. Complete backbone assignment (76 of 76 non-proline residues) was achieved using NMRViewJ aided by Probabilistic Interaction Network of Evidence algorithm (49).

Aliphatic side-chain assignments of Brd4 ET domain were obtained from 3D ^{15}N-edited TOCSY, HCCH-TOCSY, HBHA(CO)NH, and CC(CO)NH-TOCSY (44) of [U-^{13}C,^{15}N] Brd4 ET in complex with the MLV IN EBM. Side-chain ^{1}H assignments were 74% complete; most of the missing assignments were due to signal overlap of Lys side chains and nonstereospecific assignments of methylene protons. Excluding these factors the side chains were assigned to 92%.

MLV IN EBM proton chemical shift assignments in the bound form were obtained by analysis of 2D ^{13}C/^{15}N-filtered TOCSY (DIPSI2 with 60-ms mixing time), double-quantum filtered COSY, and NOESY (250-ms mixing time) (50, 51) spectra recorded on a peptide sample in complex with [U-^{13}C, ^{15}N] Brd4 ET, resulting in 88% assignment of the nonlabile protons.

Intraprotein distance restraints were obtained from 3D ^{15}N-edited NOESY and 3D ^{13}C-edited NOESY spectra recorded in water (44). Intrapeptide distance constraints were obtained from 2D ^{13}C/^{15}N-filtered NOESY (51). Intermolecular NOEs were obtained from 3D ^{13}C/^{15}N-filtered (*f1*) ^{15}N-edited (*f3*) NOESY and 3D ^{13}C/^{15}N-filtered (*f1*), ^{13}C-edited (*f3*) NOESY (44, 51). Some of the NOESY data sets (3D ^{13}C-edited NOESY, 3D ^{13}C/^{15}N-filtered (*f1*), ^{13}C-edited (*f3*) NOESY, 2D ^{13}C/^{15}N-filtered NOESY) were repeated with the lyophilized sample reconstituted in D_2O. NOESY cross-peaks were analyzed and assigned using NMRViewJ, and NOEs were categorized into strong, medium, and weak corresponding to upper restraint distances 3.5, 4.5, and 5.5 Å, respectively. NOE-derived distance constraints were supplemented with torsion angle restraints predicted by TALOS-N (52) and hydrogen-bond constraints identified in the secondary elements. A total of 150 structures were generated using CYANA 3.97 (53), from which an ensemble of the 20 structures having the lowest CYANA target function was selected. The 20-structure ensemble contained Ramachandran plot statistics of 88.7% of residues in the most favored regions, 10.1% in additionally allowed regions, 1.3% in generously allowed regions, and 0.0% in disallowed regions (Table S1).

Mass Spectrometry-Based Proteomic Analysis. To identify cellular binding partners of FLAG-Brd4(1–720) or FLAG-Brd4(1–720) mutants, nuclear extracts of HEK293T cells containing ectopically expressed FLAG constructs were bound to FLAG beads (Sigma) in 25 mM Tris (pH 8.0), 200 mM NaCl, 0.2% Nonidet P-40, 2 mM β-mercaptoethanol, and 1× complete protease mixture (Roche). Nuclear extracts were prepared using the NE-PER Nuclear and Cytoplasmic Kit (Thermo Scientific), and the extracts were allowed to bind to the beads and were subjected to an additional two washes. The bound proteins were separated by SDS/PAGE, and entire lanes were excised for in-gel trypsin digestion, followed by analysis of the peptide fragments with a hybrid LTQ Orbitrap (Thermo Scientific) equipped with a nanoelectrospray ion source (Thermo Scientific) connected to an EASY-nLC. Raw data were converted to Mascot Generic Format using ReAdW4Mascot2 (chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:start) and spectra were searched with Mascot (v.2.2.06; Matrix Science) against the UniProt human sequences with a 10-ppm precursor tolerance and 0.5-Da fragment ion tolerance. Mascot results were loaded into Scaffold (v.3.00.06; Proteome Software) with a protein probability threshold of 90%, a peptide probability threshold of 90%, and a minimum of two peptides per protein (Dataset S1). For proteins that had (*i*) no hits in the background experiments, (*ii*) three or more hits with the WT protein, and (*iii*) no hits with the mutant proteins, the number of entries in the BioGrid database (thebiogrid.org/) are indicated.

Protein Affinity Capture (Pull-Down) Experiments. Interactions between recombinant GST-MLV IN and ectopically expressed WT and mutant FLAG-Brd4(1–720) (80 μg total protein) were monitored by affinity pull-down assays as previously described for FLAG-Brd3 (6) with detection by Western blotting with anti-FLAG antibody (Abcam). Protein pull-downs were performed with glutathione Sepharose 4B beads (GE Healthcare) with buffer of 50 mM Hepes, pH 7.5, 200 mM NaCl, 0.5% Nonidet P-40, 2 mM BME, and 1× complete protease mixture (Roche). Competition between native NSD3, MLV IN EBM, and recombinant His-Brd4 ET or ectopically expressed FLAG-Brd4(1–720) was monitored by affinity pull-down using Ni Sepharose 6 fast flow beads (GE Healthcare) or anti-FLAG resin (Sigma) with detection by Western

blotting with anti-NSD3 antibody (Abcam) (12). Nuclear extracts of HEK293T cells (80 μg total protein) were prepared as described above and added to His-Brd4 ET (2 μM) with or without prebound MLV IN EBM (10 μM) or RIPA extract of HEK293T cells (100 μg) with ectopically expressed FLAG-Brd4(1–720) were incubated with or without MLV IN EBM (10 μM) before anti-FLAG pull-downs. Pull-down buffer was 50 mM Tris, pH 7.5, 200 mM NaCl, 0.5% Nonidet P-40, 2 mM BME, 20 mM imidazole, and 1× complete protease mixture (Roche) or the same buffer minus imidazole and BME, respectively.

Note Added in Proof. While this manuscript was under review, Shen et al. (54) showed that nonpolar residues in the ET domain, which we show here to be involved in binding to MLV IN (L630, I654, and F656), are also critical for binding to NSD3.

ACKNOWLEDGMENTS. We thank Matthew Plumb, Nikoloz Shkriabai, Annie Moradian, Roxana Eggleston-Rangel, and Michael Sweredoski for help with bioinformatics analysis and proteomics experiments and Punit Uphadyaya, Dehua Pei, Michael Poirier, Aparna Unnikrishnan, Claire Chandler, Chris French, and Geoff Rosenfeld for encouragement and discussions. This work was supported in part by NIH Grant AI062520 (to M.K.) and NIH Grant AI124463 (to M.P.F.). The Proteome Exploration Laboratory (S.H.) is supported by the Gordon and Betty Moore Foundation through Grant GBMF775 and the Beckman Institute.

1. Devaiah BN, Singer DS (2013) Two faces of brd4: Mitotic bookmark and transcriptional lynchpin. *Transcription* 4(1):13–17.
2. Belkina AC, Denis GV (2012) BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer* 12(7):465–477.
3. McBride AA, McPhillips MG, Oliveira JG (2004) Brd4: Tethering, segregation and beyond. *Trends Microbiol* 12(12):527–529.
4. Wu SY, Chiang CM (2007) The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem* 282(18):13141–13145.
5. Yang Z, et al. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* 19(4):535–545.
6. Sharma A, et al. (2013) BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci USA* 110(29):12036–12041.
7. Zhu J, et al. (2012) Reactivation of latent HIV-1 by inhibition of BRD4. *Cell Reports* 2(4):807–816.
8. Floyd SR, et al. (2013) The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. *Nature* 498(7453):246–250.
9. Zeng L, Zhou MM (2002) Bromodomain: An acetyl-lysine binding domain. *FEBS Lett* 513(1):124–128.
10. Brand M, et al. (2015) Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem Biol* 10(1):22–39.
11. Filippakopoulos P, Knapp S (2014) Targeting bromodomains: Epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* 13(5):337–356.
12. Larue RC, et al. (2014) Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. *Nucleic Acids Res* 42(8):4868–4881.
13. Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A (2014) Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* 42(16):10209–10225.
14. Liu W, et al. (2013) Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell* 155(7):1581–1595.
15. Rahman S, et al. (2011) The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* 31(13):2641–2652.
16. French CA, et al. (2014) NSD3-NUT fusion oncoprotein in NUT midline carcinoma: Implications for a novel oncogenic mechanism. *Cancer Discov* 4(8):928–941.
17. Ottinger M, et al. (2006) Kaposi's sarcoma-associated herpesvirus LANA-1 interacts with the short variant of BRD4 and releases cells from a BRD4- and BRD2/RING3-induced G1 cell cycle arrest. *J Virol* 80(21):10772–10786.
18. Gupta SS, et al. (2013) Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J Virol* 87(23):12721–12736.
19. De Rijck J, et al. (2013) The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Reports* 5(4):886–894.
20. LaFave MC, et al. (2014) MLV integration site selection is driven by strong enhancers and active promoters. *Nucleic Acids Res* 42(7):4257–4269.
21. Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300(5626):1749–1751.
22. De Ravin SS, et al. (2014) Enhancers are major targets for murine leukemia virus vector integration. *J Virol* 88(8):4504–4513.
23. Aiyer S, et al. (2014) Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res* 42(9):5917–5928.
24. El Ashkar S, et al. (2014) BET-independent MLV-based vectors target away from promoters and regulatory elements. *Mol Ther Nucleic Acids* 3:e179.
25. Aiyer S, et al. (2015) Structural and sequencing analysis of local target DNA recognition by MLV integrase. *Nucleic Acids Res* 43(11):5647–5663.
26. Lin YJ, et al. (2008) Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. *Protein Sci* 17(12):2174–2179.
27. Fassati A, Goff SP (1999) Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. *J Virol* 73(11):8919–8925.
28. Risco C, Menéndez-Arias L, Copeland TD, Pinto da Silva P, Oroszlan S (1995) Intracellular transport of the murine leukemia virus during acute infection of NIH 3T3 cells: Nuclear import of nucleocapsid protein and integrase. *J Cell Sci* 108(Pt 9):3039–3050.
29. Cherepanov P, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 278(1):372–381.
30. Ciuffi A, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11(12):1287–1289.
31. Llano M, et al. (2006) An essential role for LEDGF/p75 in HIV integration. *Science* 314(5798):461–464.
32. Shun MC, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 21(14):1767–1778.
33. Ferris AL, et al. (2010) Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci USA* 107(7):3135–3140.
34. Busschots K, et al. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* 280(18):17841–17847.
35. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci USA* 102(48):17308–17313.
36. Cherepanov P, et al. (2005) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol* 12(6):526–532.
37. Roth MJ (1991) Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. *J Virol* 65(4):2141–2145.
38. Xie W, et al. (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol Cell Biol* 21(19):6606–6614.
39. Brady TL, Fuerst PG, Dick RA, Schmidt C, Voytas DF (2008) Retrotransposon target site selection by imitation of a cellular protein. *Mol Cell Biol* 28(4):1230–1239.
40. Muller S, Filippakopoulos P, Knapp S (2011) Bromodomains as therapeutic targets. *Expert Rev Mol Med* 13:e29.
41. Filippakopoulos P, et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468(7327):1067–1073.
42. Hewings DS, et al. (2012) Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. *J Med Chem* 55(22):9393–9413.
43. Nicodeme E, et al. (2010) Suppression of inflammation by a synthetic histone mimic. *Nature* 468(7327):1119–1123.
44. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* 34(2):93–158.
45. Bax A, Grzesiek S (1993) Methodological advances in protein NMR. *Acc Chem Res* 26(4):131–138.
46. Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* 16(1):22–34.
47. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
48. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313–352.
49. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR (2009) Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLOS Comput Biol* 5(3):e1000307.
50. Zwahlen C, et al. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J Am Chem Soc* 119(29):6711–6721.
51. Breeze AL (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog Nucl Mag Res Sp* 36(4):323–372.
52. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213–223.
53. Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273(1):283–298.
54. Shen C, et al. (2015) NSD3-short is an adaptor protein that couples BRD4 to the CHD8 chromatin remodeler. *Mol Cell* 60(6):847–859.